



**QUEEN'S
UNIVERSITY
BELFAST**

A labor theoretic approach to information retrieval

Warner, J. (2008). A labor theoretic approach to information retrieval. *Journal of the American Society for Information Science and Technology*, 59(5), 731-741. <https://doi.org/10.1002/asi.20782>

Published in:

Journal of the American Society for Information Science and Technology

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2008 ASIS&T.

This is the peer reviewed version of this article, which has been published in final form at doi: 10.1002/asi.20782. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

This is a postprint of an article published in the Journal of the American Society for Information Science and Technology copyright © 2008 (American Society for Information Science and Technology) at <http://onlinelibrary.wiley.com/doi/10.1002/asi.20782/abstract>.

A labor theoretic approach to information retrieval

Julian Warner

**School of Management and Economics
The Queen's University of Belfast**

j.warner@qub.ac.uk

Abstract. This article synthesizes the labor theoretic approach to information retrieval. Selection power is taken as the fundamental value for information retrieval and is regarded as produced by selection labor. Selection power remains relatively constant while selection labor modulates across oral, written, and computational modes. A dynamic, stemming principally from the costs of direct human mental labor and effectively compelling the transfer of aspects of human labor to computational technology, is identified. The decision practices of major information system producers are shown to conform with the motivating forces identified in the dynamic. An enhancement of human capacities, from the increased scope of description processes, is revealed. Decision variation and decision considerations are identified. The value of the labor theoretic approach is considered in relation to pre-existing theories, real world practice, and future possibilities. Finally, the continuing intractability of information retrieval is suggested.

Introduction

Selection power, or the ability to discriminate between objects or representations of objects, has been conceived as the primary aim for information retrieval systems. Selection power is understood as a quality of human consciousness, which can be assisted or frustrated by system design, but which does not inhere in a system itself. The analogy between selection power and the radical or etymological sense of intelligence, from *inter-legere* or to choose between (Stevens, 1998, p.66), indicated that the capacity for selection was fundamental to being fully human. For information retrieval systems, selection power was produced by *selection labor*, including the direct human mental labor of cataloging, classification, and database description, and, at a further level of analysis, those aspects of human mental labor which had been transferred to information technology.

Selection labor was understood historically in relation to dominant information technologies. Orality emerging into written literacy, written literate or *premodern*, and computer-based or *modern* modes were differentiated. Selection labor had emerged as a relatively unified activity in orality emerging into written literacy and separated into description and search labor during the development of written literacy. It was reconverging as a single substantive category under modern conditions, although distinctions between description and search labor and processes remained both substantive and analytically valuable. The fundamental proposition that selection power was produced by selection labor held across orality emerging into literacy, premodernity, and modernity.

Selection labor, and its components of description and search labor, were understood as forms of mental labor. For mental labor, as for physical and productive labor, aspects of direct human labor could be transferred to humanly constructed technology, where it would be transformed into a machine process. Semantic mental labor, the labor involved in transformations motivated by the meaning of signs, remained irreducibly directly human, while syntactic labor, the labor required for pattern motivated transformations, could be delegated to technology. The transfer of syntactic mental labor to technology tended to be compelled by the lower costs of machine processes, with the discrepancy in costs between human labor and machine processes becoming increasingly acute with modern technologies.

The essential components of what can be called a labor theoretic approach to information retrieval have, then, been discursively established (Warner, 2007a, b), but a synthesis is still needed. The synthesis should reveal the economy and power of the approach, for both theoretical and practical purposes, by emphasizing the coherence and mutual relation of elements, such as selection power and selection labor, which have been previously isolated and investigated separately. The dynamism involving the transfer of human mental labor to machine processes, emerging from the costs of direct human mental labor and the common desire to avoid those costs, is then empirically illustrated by reference to major information systems. Decision practices of major real world information system are shown to embody, and to conform with, the strongly determining forces identified. The value of the labor theoretic approach will also be considered and the continuing intractability of the process of information retrieval suggested.

Method of synthesis

A full synthesis of the labor theoretic approach needs to recapitulate its discursive development and to lay bare the logical structure and progression which has lain underneath the discourse. The economy of the approach should be revealed by the limited number of concepts and activities distinguished. Its power lies in its ability to comprehend significant empirical developments in major information systems and selectively to absorb pre-existing theories. The synthesis requires explicit articulation of the underlying method of argument and the specification of the particular techniques to be employed for the full exposition of the argument, within that method.

Method

The argument in the sequence of articles has had a logical structure but not a reductionist approach. The empirical richness of information retrieval has been preserved and incorporated into a cohesive structure, discursively rather than formally presented, to this point. The underlying method has corresponded to a sophisticated understanding of the 'geometric method'. For the Italian philosopher, Giambattista Vico:

the whole secret of the geometric method comes to this: first to define the terms one has to reason with; then to set up certain common maxims agreed to

by one's companion in argument; finally, at need, to ask discretely for such concessions as the nature of things permits, in order to supply a basis for arguments, which without some such assumption could not reach their conclusions; and with these principles to proceed step by step in one's demonstrations from simpler to more complex truths, and never to affirm the complex truths without first examining singly their component parts

(Vico, 1725/1990, pp.125-126)

In accord with this view, terms defined or elucidated, a common maxim, and concessions or assumptions have been effectively, although partly implicitly, distinguished.

Theoretical and empirical analogues have been indicated for concessions requested and concessions have been rooted in ordinary discourse distinctions requested. A consistent progression from simple to more complex truths, with the components of the complex truths exemplified, was also observed. In the review and synthesis which follows, distinctions between terms defined or elucidated, the common maxim, and concessions will be made fully explicit and the economy and power of the model fully revealed.

The geometric method is known to give rise to particular difficulties, particularly when it is formally rather than discursively expressed. Characteristically, the 'demands of formal simplicity and exactitude on the one hand, and closeness of fit to reality on the other, are in conflict' (Wilson, 1973, p.470). In this instance, the geometric method has been restricted to the central concepts and the discursive presentation may also have helped preserve empirical richness. Concepts, such as selection power and selection labor, were abstracted from everyday practice, from the 'real life process' (Marx, 1858/1973, p.706), rather than theoretically imposed. Reducing the discursive development to a more formal presentation should, then, add clarity and reveal the rigor of the argument, while maintaining the strong correspondence to reality.

Technique

The geometric method is often associated with certain techniques or formalisms, such as those of symbolic logic, but need not be reduced to such techniques. In this context, the priority given to ordinary discourse concepts and real life processes over theoretical abstractions is matched by a progress in exposition from full written discourse (Warner, 2007a, b), through increasingly summary recapitulations and a relevant formalism, of formal logic, to be presented here. Intelligibility to readers familiar with ordinary discourse, but not fully conversant with the formalism, will be sustained by embedding the formalism in a summary discursive recapitulation and by complementing the formal exposition with a diagram, whose iconicity may aid intelligibility.

Elements from the formalism of formal or symbolic logic will be introduced to expose the underlying rigor of the argument. Connectives from formal logic are sufficient fully capture relations between the 'atomic fact[s]' distinguished (Wittgenstein, 1922/1981, § 2.062), with atomic facts understood to include both concepts and activities, for instance selection power as a concept and selection labor as an activity.

Synthesis

The first concept treated was *selection power*, received as an atomic fact or primitive proposition, open to elucidation but not to definition in the sense of decomposition into more primitive terms. Selection power was understood as the human faculty for discrimination which could be augmented by technologies but remained a quality of human consciousness. The first proposition, then, was to assert the value for information retrieval of

Selection power

received as a primitive proposition ¹.

Labor was regarded as comparably fundamental to selection or intelligence and as an inescapable condition of human existence. Labor was understood to include mental as well as physical labor, both as an adjunct to physical or productive labor and as an activity in itself. The specific concern was with *selection labor*, as a form of mental labor. Selection labor could then also be asserted as a primitive activity, although as something inescapably imposed by the need for selection power, rather than implicitly or explicitly sought after as a good.

Selection labor

Under a range of historical conditions, emerging with orality transforming into literacy, fully realized with literacy or premodern information technologies and continuing, although transformed, with modern technologies, selection power was produced by selection labor. A proposition to connect the primitive concept and activity was then developed, that selection power was produced by selection labor (Warner, 2007a). The proposition had empirical correlates in historical and current document, object, and people description processes. Selection labor could be decomposed, particularly clearly and strongly under premodern conditions, into description and search labor.

The production of selection power by selection labor can be regarded as analogous to material implication in formal logic, with selection power implying selection labor.

Selection power → Selection labor

Making the analogy of the production of selection power by selection labor with the relation of material implication in formal logic explicit adds value to the discursive exposition. The relation of material implication would remain valid for selection labor without selection power ². It then caters for the case, met in practice, that selection labor would not produce selection power where unhelpful or inappropriate selection labor has been applied. The first place given in material implication to selection power also suggests that selection labor can be hidden or disguised from view behind selection power, for instance, congealed in the products of selection labor. The disguising of

selection labor would also correspond to its theoretical neglect, despite its costs and significance (Hayes, 2000), and may be partly responsible for that neglect. In the practical use of a particular system, a searcher might have to determine the nature of selection labor or process applied from the selection power afforded: for instance, variant and unreconciled forms of a single author's name could be taken to indicate the absence of effective human description labor or machine processes in reconciling variants.

Some further propositions, corresponding to 'such concessions as the nature of things permits', in Vico's account of the geometric method, which had analogues in ordinary discourse and common understandings, were then introduced to illuminate the relations of selection labor to description and search labor and the possibilities of their transfer to technology. Concepts isolated and distinctions made corresponded to real world practice, suggestive of their robustness.

Labor was understood to include mental labor.

Labor includes mental labor

Mental labor shared some characteristics with physical and productive labor. Direct human mental labor could be transferred to technology, with technology itself regarded as a human construction.

Human mental labor can be transferred to Information technology

When fully transferred human mental labor becomes a machine process, in the intervals between direct human intervention.

Some categories developed primarily for physical and productive labor did explicitly recognize elements of mental labor. Universal labor, 'all scientific work, all discovery and invention', was differentiated from communal labor: universal labor is 'brought about partly by the cooperation of men now living, but partly also by building on earlier work', while communal labor 'simply involves the direct cooperation of individuals' (Marx, 1894/1981, p.199). Labor might originate as communal labor, but, as the knowledge created by labor diffuses, it is transformed into universal labor.

Labor separates into Communal labor and Universal labor

Information technology, in both its hardware, and increasingly, its software aspects, would be regarded as the product primarily of universal labor. Human descriptions of information objects, such as catalog records, would be the products of predominantly communal labor.

Distinctions of labor, process, and product, understood close to their ordinary discourse senses, could also be applied to mental labor and its products. With oral speech, particularly under orality prior to the development of written language, labor includes the whole activity of communication, with process and product not fully separated out. With

written language, there is a separation of product from labor and the possibility of formalizing the processes involved in the making of that product. Under modernity, processes can be automatically conducted. An originally undifferentiated labor, then, separates into labor, process, and product, with associated changes in the senses of the constituent terms, or:

Labor separates into Labor, process, and product

Analytic distinctions between labor, process, and product can then be applied to modern information retrieval activities and systems, with labor as human labor, process as a formalized process, often delegated to technology, and product as the outcome of labor and process, for instance, records or index descriptions.

Distinctions specific to mental labor were then introduced. Mental labor could be semantic or syntactic in character, linking a widely held distinction between levels of analysis to mental labor. Semantic and syntactic mental labor could be regarded as separating out from undifferentiated mental labor.

Human mental labor separates into Semantic labor and Syntactic labor

An analytic distinction could then be made between semantic and syntactic mental labor, for both premodernity and modernity. Semantic mental labor corresponded to the work involved in conducting transformations on signs motivated by their meaning, and syntactic labor to transformations reduced to motivation from patterns. Semantic mental labor was irreducibly directly human

Semantic labor → Human labor

Expressing the relation between semantic labor and human labor as material implication allows for the possibility of human labor being more extensive than semantic labor, incorporating syntactic aspects³. In contrast to semantic labor, syntactic labor could be transferred to technology, with labor becoming process.

Syntactic labor can be transferred to Information technology (labor becomes process)

A technological process was *a priori* syntactic in character, operating on patterns not directly on meaning. In premodern historical practice, syntactic labor had been delegated to human clerical labor, but, under modern conditions, was increasingly being transferred to information technology as a machine process.

Synthesis of maxims and concessions

Selection power remained a relatively stable primitive, as a quality of human consciousness, while selection labor, although persistently present, modulated with historical transformations in information technology. Our primary attention, then, in

synthesizing the maxim that selection power is produced by selection labor with the distinctions established within labor, must be to the transformations of selection labor, as a form of mental labor.

Under orality emerging into literacy, description labor in information systems included the cognitive labor of memory and recall and the bodily and communicative labor of public oral expression (the Icelandic law speaker was given as an example of such an information system, embodied in socially designated individual) (Njal, 1280/1960, pp.306-308). With oral speech, the process of production and the product, as audible speech, disappeared with the process, leaving no trace outside the memory of the auditors. Process and product were not fully separated from labor. A distinction of syntactics from semantics could be recovered, but is not usually strongly marked in oral societies.

For information systems under written literacy or *premodernity*, selection labor progressively separates into description and search labor, with search labor emerging as the work involved in traversing the products of description labor.

Selection labor *separates into* Description labor and Search labor

Analytically and substantively, selection labor can be regarded as the sum of description and search labor, once these aspects have separated out. The historical separation out of description and search labor from selection labor supports the proposition that selection labor can be distributed between description and searching, but that its overall quantity cannot be diminished, below certain limits associated with the number and variety of objects for selection. The separation out of products from labor introduces the possibility of the distribution of the products of description labor, such as catalog records, catalogs, and bibliographies, with human labor congealed in these products. Aspects of processes of description can also be delegated to clerical human labor, assisted by the technologies associated with writing and by special purpose information machines for the production, and, at a later date, sorting of written utterances.

Under *modernity*, for both description and search labor, the distinction between semantic and syntactic labor is sharpened, by the possibility and actuality of the transfer of syntactic labor to technology.

Description labor *separates into* Description labor semantic and Description labor syntactic

Search labor *separates into* Search labor semantic and Search labor syntactic

The semantic aspects of description and search labor, which include the understanding of syntactic processes, remain irreducibly directly human.

Description labor semantic → Human labor

Search labor semantic → Human labor

The syntactic aspects of description and search processes, previously conducted by direct human work, can be transferred to information technologies, predominantly to the computer as a universal information machine programmed to function as an appropriate special purpose information machine.

Description labor syntactic *can be transferred to* Information technology (labor becomes process)

Search labor syntactic *can be transferred to* Information technology (labor becomes process)

Description and search labor are, then, more fully revealed as both semantic and syntactic in character.

Summary

The synthesis of maxims and concessions has, then, revealed a pattern of progressive development, with originally undifferentiated categories, such as labor and selection labor, separating out into different aspects.

The central propositions underlying the discursive exposition can also be presented as a sequence of summary statements, with the distinctions between primitive terms, maxim, concessions requested, and the synthesis of maxims and concessions, retained (See Figure 1). The order of the main section of the discursive developments of the labor theoretic approach, both in the full expositions in the original articles (Warner, 2007a, b) and in the recapitulation here, corresponds to the sequence of statements and could be mapped to it. The logical connective of material implication is used to relate concepts and activities distinguished where it adds value to understandings obtainable from ordinary discourse. Material implication was classically the most difficult and productive of the logical connectives and has here emerged as a highly significant crux, particularly illuminating the production of selection power by selection labor.

Central propositions can also be presented as a diagram, whose iconicity may aid intelligibility (See Figure 2). The representation given in the diagram is restricted to the primitive terms, common maxim, and synthesis, with concessions implicitly incorporated but not explicitly shown. The diagram can be read iconically in a number of complementary ways. Selection power is given priority by being placed at the head of the diagram, resting on selection labor, whose submerged position is congruent with its congealing and disguise in its products, with its theoretical neglect, and with its significance, when exposed, in supporting selection power. If the line indicating the distribution of selection labor between description and searching is also read as the beam of a scales or balance, it must be considered as a hinged or potentially broken beam. For instance, the quantity of description labor could be increased without reducing search labor, if that description labor is not helpful to the searcher. The analogy with the beam

of a scales remains valuable for giving iconic form to the idea that labor can be distributed between description and searching, but cannot be reduced below a given quantity of selection labor.

In its current development, the argument is slightly static, with no dynamic compelling the transformation of relations between categories from *can be* to *is*.

Transformation into a dynamic

A dynamism which compels the transfer of human syntactic labor to technology can be found in the conditions of direct human labor, including its costs and elements of drudgery⁴. Discerning a common preference, particularly under capitalism, for a reduction in the costs of processes, by abridging direct human labor, is a specific, and historically rooted, position, distinguishable from Zipf's universalized principle of least effort or of resistance to labor (Zipf, 1936). The equivalence of the automatic machine to slave labor (Wiener, 1954, p.162), a condition not often voluntarily adopted, should also be recalled. *Can be* relations between categories are transformed into *is*.

Syntactic labor *can be transferred to* Information technology (labor becomes process)

is transformed into

Syntactic labor *is transferred to* Information technology (labor becomes process).

The transfer would apply to both syntactic description and to syntactic search labor.

Description labor syntactic *is transferred to* Information technology (labor becomes process)

Search labor syntactic *is transferred to* Information technology labor (labor becomes process).

Direct human labor then becomes primarily semantic labor, with the transformation similarly applying to both description and search labor, aspects of selection labor.

The concepts, of selection power and labor, and of semantic and syntactic labor and processes, which form components of the model developed, have been abstracted out from historical and current information practice, and from ordinary discourse, rather than imposed. In such a process of abstraction, there is a partial analogy with the construction of models of the computational process, not as ends in themselves, but as aids to understanding. Once fundamental matters are grasped, we can bring back to the practical world this understanding, 'which we could never obtain while immersed in inessential detail and distraction' (Minsky, 1967, pp.2-3). Automata theory has tended to be develop models of the computation process partly as ends in themselves, becoming increasingly

technical (Boolos and Jeffrey, 1989), rather than as an endeavor which might inform practical understanding. The challenge here is to show that the determining forces identified for information systems can be discovered as influential for real world systems.

Decision practice

Decision practice

Major information systems producers would not formulate their policies and activities in terms of the concepts identified here, but their activities can be seen to be constrained and guided by the determining forces identified, across a range of significant information systems. The effect of the determining forces can vary with the intentions and market of the producer

A common, and not greatly varying, characteristic of the movements produced by the motivating forces would be the transfer of direct human labor to technological processes. Description processes, such as the creation of indexes from records, which, under premodern conditions, were carried out by human syntactic labor have been largely transferred to technology, strongly implying a desire to avoid the costs of direct human labor. Description processes, which might have been used under premodernity to create certain description products in fixed media, for instance, different orders for indexes or records, are effectively moved from description to potential instantiation in searching, economizing on the creation of products and enabling a greater variety of orders.

Systems with syntactically based description processes, with human labor transferred to technology, in contrast to premodern practices, have proliferated, exemplified by Internet search engines. Primarily syntactic and machine processes are increasingly used to generate descriptions which can be used to search the texts of works or citations between works, for both *Amazon.com* (Amazon, 2007) and *Google* (2007a,b,c). The human labor embodied in the written text of documents described is thereby taken as a resource for description. These description processes can aid specificity in retrieval, but the very number and diversity of results obtained may tend to transfer selection labor to the searcher, if they are used. The links, or traces left by semantically guided explorations of resources, for instance, may be exploited for the order of references in retrieval and could be regarded as products of a form of unpaid semantic description labor. The increasing dominance of *Google* as a search engine which might be taken to contradict the assertion of the proliferation of syntactically based systems, can be regarded as the diffusion of an effective syntactic system, with some monopolistic forces aiding its dominance. The market for information services has been regarded as acting as an agent of error elimination or correctability, as a regulative mechanism, with a better mechanism of correctability yet to be demonstrated (Swanson, 1980, p.128). Although syntactic processes for selection and ordering of documents may be open to unlimited variations derived from the primitive computational operations, only some of these variations may be interesting or useful. One task for information science would be to attempt to understand the effectiveness of *Google*, including the effects of syntactic transformations on the semantic interpretation of words and phrases.

An enhancement of human capacities can be derived from the greatly increased scope of syntactically generated descriptions, if these are appropriately exploited in searching. For instance, a wide range of discussions of Marx's conception of technology as, '*organs of the human brain, created by the human hand: the power of knowledge, objectified*' (Marx, 154/1973, p.706), can be discovered from *Google Advanced Book Search* by searching for the exact phrase, *power of knowledge objectified* (Google, 2007c). Such a search requires understanding of patterns of diffusion within the primary literature and of the process of searching. With regard to the primary literature, the phrase comes from the dominant English language translation of the *Grundrisse*, published in 1973, and not yet subject to processes of secondary diffusion which might have enabled discussion of the theme, from Marx, of technology as a human construction, without use of the significant phrase. For searching, the understanding that phrase searching often exclusively yields tokens of the intended type could have been experientially obtained by reiterated searches on different phrases⁵. The correspondence of tokens to a type is also theoretically explicable from the perspective of information theory in terms of the transition probabilities between letters and between words, understood as cohesive groups of letters (Shannon, 1951/1993). Semantic description processes are bypassed in the discovery of documents but may still be exploited for obtaining copies of those documents from libraries or bookshops, using their directly humanly and semantically constructed records. The enhancement of human capacities itself creates partly novel difficulties: of selection from an enlarged body of potentially relevant and now more readily discoverable primary literature, and of how to exploit syntactically generated descriptions. Search labor, as a component of selection labor, has, then, been increased and slightly transformed, following developments in description practices.

The distribution of the products of semantic description labor, such as catalog records, emerged in premodernity and continues as a practice under modernity, enhancing selection power in particular ways, with elements of continuity and modulation in the practice. The costs of the human semantic labor required for the production of descriptions are thereby effectively shared, although the semantic labor itself is not directly divided. *WorldCat*, for instance, contains the products of semantic description labor, as catalog records. Records are created by communal labor, variously empirically distributed, guided by the universal labor embodied in codes, understandings, and precedents for document description. The products of communal labor, as catalog records, are distributed to participating libraries, for syntactic and primarily technology based, incorporation into their own institutional catalogues (WorldCat, 2007). *Amazon.com* also makes use of the products of human semantic description labor, again variously empirically distributed, as records for books, although to less full and exacting standards than that required for *WorldCat* (WorldCat, 2007; Amazon, 2007). In contrast to the specificity obtainable from syntactically generated descriptions, deliberately humanly assigned descriptions may offer generic power and control. For some systems, exemplified by Internet search engines, semantic description labor, if it has been used in the assignment of metadata, cannot necessarily be directly and separately exploited in searching. A continuity with premodernity is revealed in the distribution of description products and modulation emerges in the supplanting of clerical labor by machine

processes for the integration of products into intra-institutional catalogs. The set of practices can be characterized as the distribution of the products of human semantic description labor, in order to share the costs of that labor, and the dominance of *WorldCat* understood in these economic terms.

The inheritance of the absence of precise coding for oral speech and non-written graphic forms implies a limitation on the effective extension of established syntactic procedures to such forms. *Google Image Search* relies for searching and retrieval, although not for the inspection of results, primarily on verbal descriptions generated from written verbal forms discovered in proximity to the images (Google, 2007b). Other characteristics of computer-held images, such as type of file, amenable to automatic exploitation can also be searched. Retrieved results are presented iconically, with scanning for further selection enabled by the reduced size of the images displayed. The contrast between the treatment of symbolic and iconic graphic signs implies a connection between labor and epistemology, with written language embodying accumulated or universal human encoding labor, for instance, as standardized orthography and clearly marked boundaries between words.

The determining forces isolated by the theoretical discussion have, then, emerged as highly significant influences on real world developments in major information systems. Technological processes are *a priori* syntactic in character, although this is not a recent development. The most compelling aspect of the dynamic observed, determined by relative costs, has been the transfer of syntactic processes from direct human labor to the humanly constructed and adopted technology. Human semantic description labor is retained, although not necessarily widely applied, both as an inheritance and for its current value, although inheritance and value have not necessarily been fully separated out. Where semantic description labor had value in enhancing selection power and reducing search labor, its absence will tend to transfer selection labor to the searcher. In other contexts, it has been suggested that a theory of bibliographic searching will have to replace a theory of bibliographic description (Wilson, 2001). A task for information science might be to understand the expertise and strategies which can assist the exploitation of syntactically generated descriptions. Strategies may not be complex, once they are understood: consider, for instance, the contrasting results characteristically obtained from word and phrase searching.

There is no intrinsic dynamism to the forces identified in the model articulated which compels a transfer of human semantic selection labor from description to searching, but the realization of the model, in many contexts, may have that effect. With a relatively closed and controlled system, the fact of consumption of material may be assured and patterns of consumption, one epistemological basis for the choice of mode of description (Shera, 1952/1965; 1961), can be anticipated. Even there the assignment of objects to categories can be problematic and is known to be highly inconsistent. With more open systems, the fact of consumption cannot be assured and patterns of consumption are difficult to anticipate – ‘if you watch it [display of *Google* queries] long enough, the different queries show how diverse the world is’ (Weisman, 2002). Practices which reduce human labor in description and transfer work to searching may then proliferate.

Decision variation

Decision practices reveal evident commonalities, variation, and an underlying value. An evident commonality lies in the transfer of established description processes to technology. Variation occurs primarily in two loci, in the extent of syntactic description processes applied – what proportion of the original document to capture and represent – and in the presence or absence of the products of semantic description labor, and, where present, in the type of product. A degree of human semantic search labor has emerged as inescapable. An underlying value, embodied in real world systems and influencing practice, is of selection power.

The different aspects of variation have different costs and effects. Expanding the scope of syntactic description processes, although not costless, has limited costs and may yield an enhancement of human capacities. Semantic description labor, by contrast, is costly. The presence of the products of semantic description labor may enhance selection power and reduce search labor, while their absence can be associated with limited selection power and high levels of search labor.

Decision considerations

Decision considerations can be constructed from the contrasting costs and effects of the two aspects of variation. The low costs of syntactic description processes, and the possible enhancement of human capacities obtainable, favor their instantiation. Restrictions on the scope of machine description process may only be a transitional stage, with developments inhibited by inherited beliefs and attitudes, particularly in library, rather than broader information, contexts. The primary decision consideration, which may continue and may not be a transitional phase, is to do with semantic description labor and the associated distribution of human semantic labor between description and searching. Description labor is more amenable to deliberate control than search labor the specific consideration would be, whether to apply semantic description labor, then, what type, including exhaustivity, of description product to create, and, whether to draw upon any existing products of semantic description labor. Decision considerations, then, relate primarily directly to human labor, secondarily to machine processes, and only slightly to the materials consumed in the products of labor and processes.

Value of a labor theoretic approach

A historical dynamism for information retrieval practice was identified, with primarily oral, written literate and premodern, and modern, involving computer technologies, modes of information distinguished. Selection power was relatively consistently valued, with the exception of the recently dominant, but currently eroding, information retrieval research tradition. Selection labor, by contrast, emerged experientially and as an empirically observable practice, with written literacy, then further divided into description and search labor, and was tending to reconverge as single substantive category. The validity of the proposition that selection power is produced by selection

labor was reinforced by historical and current empirical considerations. The value of the labor theoretic approach can be further considered with regard to its formal qualities as a theory, its relation to the objects described, its connection to ordinary discourse and to common experience, and for its absorption of pre-existing theories.

Formally, the theory has simplicity and economy, reducible to a brief series of statements, qualities usually strongly valued in a theory. In relation to objects described, an expansive understanding of information retrieval systems was adopted, accompanied by ostensive exemplification rather than restrictive definition. The theory was able to comprehend systems in oral, written or *premodern*, and computational or *modern*, modes. A dynamic of change and possibilities for deliberate and informed intervention within that dynamic were identified for systems under modernity. The theory can then be regarded as comprehensive and powerful in relation to the activities it takes as its object. The combination of simplicity and economy with comprehensiveness and power makes it parsimonious, possibly representing a final reduction to essential and inescapable elements.

The comprehension of real world practice is matched by a closeness to ordinary discourse and common experience. The concepts of selection power and selection labor, and the distinction of semantic from syntactic labor, have analogues in ordinary discourse conceptions and everyday information practices. The analytical distinctions introduced and employed, for instance, between labor, process, and product, all emerge from real historical transformations of lived categories. Rooting in ordinary experience should give the theory further qualities of robustness and wide applicability.

Pre-existing theories, from librarianship and indexing and classic information retrieval research, and a concept from information society discussions, have been selectively absorbed. Selection power was taken from librarianship and indexing, but the preference of those activities and disciplines for human description was transformed into a fuller understanding of the distribution of human semantic labor between description and searching, and the recognition of the possibility of transfer of forms of labor to technology. Query transformation, implicitly valued in classical information retrieval research, was incorporated into selection power, both practically and theoretically, as a special case, similarly produced by selection labor. Informational labor, recognized in information society discussion, was differentiated into semantic and syntactic mental labor, with human labor distinguished from machine processes. Technology has not been repressed, only covertly to reemerge and distort understandings, but has been explicitly incorporated into the theory. A synthesis of pre-existing theories, adopting elements of value and discarding those aspects which have obstructed understanding, has, then, been achieved.

The relation of the labor theoretic approach to pre-existing theories has a partial, and revealing, analogy to revolutions in mathematics. Classically, revolutions in mathematics have changed the interpretation, or verbal metalanguage, while retaining the form, or symbolic object-language, economizing on the accumulated intellectual labor, or universal labor, embodied in known transformations in the object-language:

And in thus preserving the form while modifying the interpretation, I am following the great school of mathematical logicians who, in virtue of a series of startling definition, have saved mathematics from the sceptics, and provided a rigid demonstration of its propositions.

(Ramsey, 1925/1990, p.219)

In this instance, the interpretation has been brought into rigorous accord with essential or inescapable components of the form or practice, particularly through the idea of selection. The selection inherent in the process of retrieval is matched by the value placed upon selection, with the addition of power, as the fundamental value for information retrieval. The *value* of selection power is then congruent with the *process* of selection.

The theory then meets a rigorous test for knowledge, as ‘an ideal reproduction of the external world serviceable for cooperative action thereon’ (Childe, 1956, p.54). The theory has elements of idealization in its abstraction out of common activities from empirical reality, but the abstraction then gives an enhanced understanding of those activities, for instance in the distinction between semantic and syntactic mental labor and their relation to technology. Distinctions established have been returned to the external world, yielding an enhanced understanding of it and identifying a dynamic for change. The theory is also serviceable for cooperative action, enabling guided and deliberate intervention in the dynamic, for instance with regard to the distribution of semantic labor between description and searching, rather than simply a not fully conscious reproduction of the dynamic.

The owl of Minerva takes flight at evening, or understanding grows towards the end of a process, and, in this instance, human description labor has been fully identified as a category as it is being complemented, and challenged, by syntactic machine description processes. For future development, particular aspects of the labor theoretic approach may obtain special significance. For instance, the enhancement of human capacities enabled by the increased scope of description processes also raises the issue of how most effectively to exploit the descriptions created. Existing research themes could be transformed and carried forward: for instance, the recognition of the multi-faceted nature of relevance, previously understood as requiring different methods of human description (Wilson, 1973), could now be conceived as how to exploit syntactically generated descriptions along the dimensions of relevance identified. Fuller empirical richness would be possible within the conceptual framework established.

The gestalt of the computer (Rosenberg, 1974), in the particular sense of extravagant expectations of the computational process and of the transfer of human intelligence to technology, which may have been one motivating factor in the growth and diffusion of the classical information retrieval tradition, with its implicit transfer of human judgment to technological processes in query transformation, is dissolving, with the increased diffusion of computational technologies. The contrast between mystical practice, in the

sense of proposals for systems, and restrained theory, with the theory of computation coinciding with ordinary discourse conceptions of computability, is being reduced by the experience of the practice of computation. The revolution in the mechanization of mental labor (Minsky, 1967, p.2) can now be seen as a late 20th century revolution, currently stabilizing at the syntactic level of computational transformations and the interface to systems.

Conclusion

At the semantic level, the process of information retrieval may remain far more enduringly intractable, not amenable to teleological transformation, despite technological and system developments, than has been generally conceded. One mid-18th century comment, produced by a lexicographer who relied partly on his own accumulated knowledge for making a monolingual dictionary (Boswell, 1791/1980, pp.131-132) ⁶, still gives the most convincing account of the process of information retrieval.

When first I engaged in this work, I resolved to leave neither words nor things unexamined, and pleased myself with a prospect of the hours which I should revel away in feasts of literature, the obscure recesses of northern learning, which I should enter and ransack, the treasures with which I expected every search into those neglected mines to reward my labour, and the triumph with which I should display my acquisitions to mankind. ... But these were the dreams of a poet doomed at last to wake a lexicographer. I soon found that it is too late to look for instruments, when the work calls for execution, and that whatever abilities I had brought to my task, with those I must finally perform it. To deliberate whenever I doubted, to enquire whenever I was ignorant, would have protracted the undertaking without end, and, perhaps, without much improvement; for I did not find by my first experiments, that what I had not of my own was easily to be obtained: I saw that one enquiry only gave occasion to another, that book referred to book, that to search was not always to find, and to find was not always to be informed; and that thus to pursue perfection, was, like the first inhabitants of Arcadia, to chase the sun, which, when they had reached the hill where he seemed to rest, was still beheld at the same distance from them.

I then contracted my design, determining to confide in myself, and no longer to solicit auxiliaries, which produced more incumbrance than assistance: by this I obtained at least one advantage, that I set limits to my work, which would in time be finished, though not completed.

(Johnson, 1755/1982, pp.21-22)

Mental labor is both made explicit and also alluded to, in the repeated analogies made between literature searching and physical mining. The Aristotelian notion of deliberation – ‘to deliberate whenever I doubted’ – is also implicitly critiqued. The contrast between ‘finished’ and ‘completed’, a subtle distinction between the meanings of words made in a

lexicographic context, has relevance here: the particular sequence of articles is ended, but the research agenda not completed.

References

- Amazon. (2007). Amazon.com. At http://www.amazon.com/ref=topnav_gw_b/102-2282358-0503334 (June 27, 2007).
- Boolos, G.S. and Jeffrey, R.C. (1989). Computability and logic. 3rd edition. Cambridge, England etc.: Cambridge University Press.
- Boswell, J. (1791/1980). Life of Johnson. Edited by R.W. Chapman and revised by J.D. Fleeman. Introduction by P. Rogers. Oxford etc.: Oxford University Press.
- Childe, V.G. (1956). Society and knowledge. London: George Allen & Unwin.
- Google. (2007a). Google. At http://www.google.co.uk/advanced_search?hl=en (June 27, 2007).
- Google (2007b). Google Advanced Image Search. http://images.google.com/advanced_image_search?hl=en (June 27, 2007).
- Google. (2007c). Google Book Search. http://books.google.co.uk/advanced_book_search (July 25, 2007).
- Hayes, R.M. (2000). Assessing the value of a database company. In B. Cronin and H.B. Atkins (Eds.), The web of knowledge: a festschrift in honor of Eugene Garfield (pp.73-84). Medford, NJ, 2000: Information Today, Inc..
- Johnson, S. (1755/1982). Preface to the Dictionary. In E.L. McAdam and G. Milne (Eds.). Johnson's dictionary: a modern selection (pp. 3-29). London and Basingstoke: Macmillan.
- Marx, K. (1858/1973). Grundrisse: foundations of the critique of political economy (Rough draft). Translated with a Foreword by Martin Nicolaus. London etc.: Penguin Books in association with New Left Review.
- Marx, K. (1894/1981). Capital: a critique of political economy. Volume Three. Introduced by E. Mandel. Translated by D. Fernbach. Harmondsworth etc.: Penguin Books in association with New Left Review.
- Minsky, M. (1967). Computation: finite and infinite machines. Englewood Cliffs, NJ: Prentice-Hall.
- Njal. (1280/1960). Njal's Saga. Translated and introduced by M. Magnusson and H. Pálsson. Harmondsworth etc.: Penguin Books.

Ramsey, F.P. (1925/1990). The foundations of mathematics. In F.P. Ramsey. Philosophical papers (pp.164-224). Edited by D.H. Mellor. Cambridge University Press, Cambridge etc..

Rosenberg, V. (1974). The scientific premises of information science. Journal of the American Society for Information Science. 25, 263-269.

Shannon, C.E. (1951/1993). Prediction and entropy of printed English. In C.E. Shannon. Collected papers (pp.194-208). Edited by N.J.A. Sloane & A.D. Wyner. Piscataway, NJ: IEEE Press.

Shera, J.H. (1952/1965). Foundations of a theory of bibliography. In J.H. Shera. Libraries and the organization of knowledge (pp.18-33). Edited by D.J. Foskett, 1965. Hamden, Connecticut: Archon Books.

Shera, J.H. (1961). Social epistemology, general semantics, and librarianship. In J.H. Shera. Libraries and the organization of knowledge (pp.12-17). Edited by D.J. Foskett, 1965. Hamden, Connecticut: Archon Books

Swanson, D.R. (1980). Libraries and the growth of knowledge. In D.R. Swanson (Ed.). The role of libraries in the growth of knowledge (pp.112-136). Chicago: University of Chicago Press.

Stevens, A. (1998). Ariadne's clue: a guide to the symbols of mankind. London etc.: Allen Lane, The Penguin Press.

Vico, G. (1725/1990). The autobiography of Giambattista Vico. Translated from the Italian by M.H. Fisch and T.G. Bergin. Ithaca and London: Cornell University Press.

Vico, G. (1710/1988). On the most ancient wisdom of the Italians: Unearthed from the origins of the Latin language: including the disputation with the Giornale de' Letterati d'Italia. First published 1710. Translated and introduced by L.M. Palmer. Ithaca and London: Cornell University Press.

Warner, J. (2007a). Selection power and selection labor for information retrieval. Journal of the American Society for Information Science and Technology. 58, 915-923.

Warner, J. (2007b). Description and search labor for information retrieval. Journal of the American Society for Information Science and Technology. 58, In press.

Weisman, R. (2002). Talking search technology: Eric Schmidt, Chairman and CEO, Google Inc. Boston Globe. April 4, 2002, p.D1.

Wiener, N. (1954). The human use of human beings: cybernetics and society. Revised edition. New York: De Capo Press.

Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*. 9, 457-471.

Wittgenstein, L. (1922/1981). *Tractatus Logico-Philosophicus*. London and New York: Routledge and Kegan Paul.

Wilson, P. (2001). Review of E. Svenonius. *The intellectual foundations of information organization*. *College and Research Libraries*. 62, 203-204.

WorldCat. (2007). WorldCat. OCLC. Retrieved June 27, 2007, from <http://firstsearch.uk.oclc.org/WebZ/FSPage?pagename=advanced:sessionid=fsapp7-45506-f3fqgebu-7vzfle:entitypagenum=3:0>.

Zipf, G.K. (1936). *The psycho-biology of language: an introduction to dynamic philology*. London: George Routledge.

Acknowledgments

I would like to acknowledge the John Campbell Trust who supported a presentation of the themes of this article at the Annual Meeting of the American Society for Information Science and Technology Annual Meeting, Providence, Rhode Island, November 2004. and the hospitality and assistance of the Research Centre for the Social Sciences (RCSS), University of Edinburgh, where the article was originally written during study leave from the Queen's University of Belfast, February–July 2005.

I would also like to record my indebtedness to the unnumbered students who interactively contributed to the developments of the research themes expounded, by taking courses in Information Policy, 2001-2006, and to the referees for their helpful and constructive comments on this and the preceding articles (Warner, 2007a, b).

Notes

1. The deliberate omission of the assertion sign, ‘ \vdash ’, which can be read verbally as, ‘It is true that ...’, is in accord with the redundancy theory of truth, traceable to at least the 17th century and Honoré Fabri (Vico, 1710/1988, pp.145-146), and receiving clear modern expression by Wittgenstein:

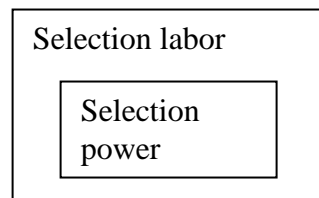
Frege’s assertion sign ‘ \vdash ’ is logically altogether meaningless; in Frege (and Russell) it only shows that these authors hold as true the propositions marked in this way.

‘ \vdash ’ belongs therefore to the propositions no more than does the number of the proposition. A proposition cannot possibly assert of itself that it is true.

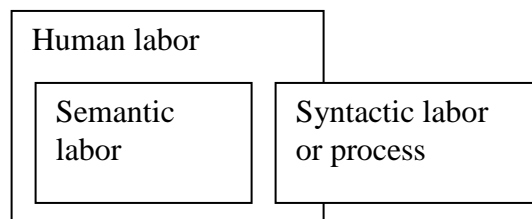
(Wittgenstein, 1922/1981, § 4.442)

Denying the possibility of a proposition asserting its own truth is consistent with a view of logic and of technology as human constructions rather than autonomous existents and with the emphasis on selection power as a property of human consciousness, assisted by, but not inhering in, system design.

2. Selection power \rightarrow Selection labor would be valid, in truth-table terms, where Selection power is False and Selection labor is True, or NOT-Selection power AND Selection labor, or, \neg Selection power \wedge Selection labor. Diagrammatically, the relation between Selection labor and Selection power could be represented as:



3. Semantic labor \rightarrow Human labor is true for NOT-Semantic labor AND Human labor, but the number of instances of this could be expected to be reduced, with modern technologies, and the quantity of human labor which is not semantic labor accordingly diminished. Diagrammatically, the relation between Human labor, Semantic labor, and Syntactic labor or process could be represented as:



4. The possibility of significantly differing costs or wage rates for comparable forms of direct human labor, particular between different geopolitical regions, is acknowledged, but not directly addressed.
5. 34 tokens of the intended type, or reference to the passage in the *Grundrisse*, were recalled, with no unintended recalled descriptions (Google, 2007c).
6. Johnson did visit libraries at Oxford as part of his work on the *Dictionary of the English Language*. Some 20 years after the publication of that work, Johnson also wrote an entry for the title-page of another work attributed to him, *An account of an attempt to ascertain the longitude at sea, by an exact theory of the variation of the magnetical needle*, in ‘the great Catalogue ... with his own hand’ (Boswell, 1791/1980, pp.190, 194).

Primitive terms

- **Selection power**
- **Selection labor**

Common maxims

- **Selection power → Selection labor**

Concessions

- **Labor includes mental labor**
- **Human mental labor *can be transferred to* Information technology**
- **Labor *separates into* Communal labor and Universal labor**
- **Labor *separates into* Labor, process, and product**
- **Human mental labor *separates into* Semantic labor and Syntactic labor**
- **Semantic labor → Human labor**
- **Syntactic labor *can be transferred to* Information Technology (labor becomes process)**

Synthesis

- **Selection labor *separates into* Description labor and Search labor**
- **Description labor *separates into* Description labor semantic and Description labor syntactic**
- **Description labor semantic → Human labor**
- **Description labor syntactic *can be transferred to* Information Technology (labor becomes process)**
- **Search labor *separates into* Search labor semantic and Search labor syntactic**
- **Search labor semantic → Human labor**
- **Search labor syntactic *can be transferred to* Information Technology (labor becomes process)**

Figure 1. Summary of synthesis.

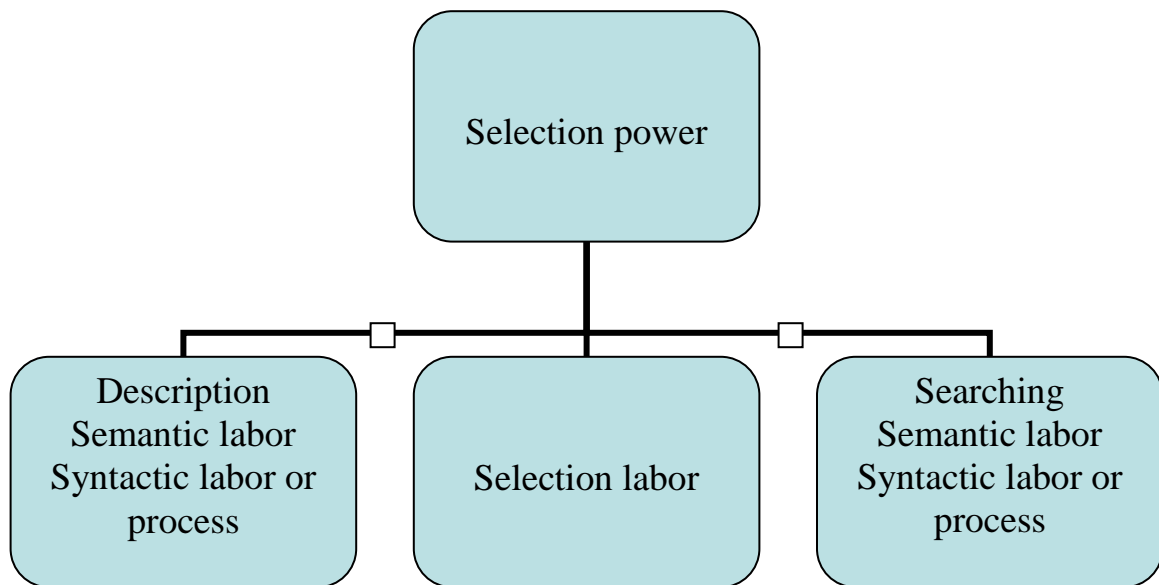


Figure 2. Representation of synthesis